PATENT ABSTRACTS OF JAPAN

(11)Publication number : **07-056595**
(43)Date of publication of application : **03.03.1995**

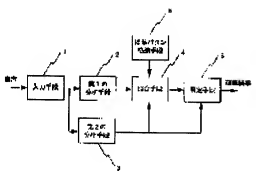
(51)Int.Cl. **G10L 3/00**
G10L 3/00

(21)Application number : **05-204915**
(22)Date of filing : **19.08.1993**

(71)Applicant : **HITACHI LTD**
(72)Inventor : **ODAKA TOSHIYUKI**
AMANO AKIO

(54) VOICE RECOGNITION DEVICE

(57)Abstract:
PURPOSE: To recognize voice while coping with the varying conditions of a user's uttering and the changes in the user by controlling a collating means or a discriminating means based on the detection results of the changes in the uttering conditions and the changes in the user.
CONSTITUTION: Voices, that are inputted and digitized through an input means 1, are acoustically analyzed for every constant time interval by a first analysis means 2 and the result of the analysis is outputted in a form which is suitable to a collating means 4. The means 4 performs collation between time sequential patterns and a standard pattern and outputs the score against each standard pattern. The score outputted from the means 4 is inputted to a discrimination means 5 and a candidate corresponding to a best scored standard pattern or plural higher ranking candidates are outputted as recognition results. A second analysis means 3 analyzes voices inputted through the means 1, extracts the changes in uttering conditions and the changes in uttering speeds, outputs these information and controls the means 4 and 5 based on the outputs.



LEGAL STATUS

[Date of request for examination]

[Date of sending the examiner's decision of rejection]

[Kind of final disposal of application other than the examiner's decision of rejection or application converted registration]

[Date of final disposal for application]

[Patent number]

[Date of registration]

[Number of appeal against examiner's decision of rejection]

[Date of requesting appeal against examiner's decision of rejection]

[Date of extinction of right]

(19)日本国特許庁（ＪＰ）　(12)公開特許公報（Ａ）　(11)特許出願公開番号

特開平7－56595

(43)公開日　平成７年(1995)３月３日

(51)Int.Cl.[6]	識別記号	庁内整理番号	ＦＩ	技術表示箇所
Ｇ１０Ｌ　3/00	５７１　Ｊ	9379－5H		
	５３１　Ｋ	9379－5H		

審査請求　未請求　請求項の数９　ＯＬ　（全７頁）

(21)出願番号　特願平5－204915

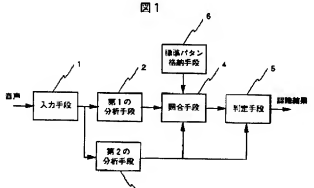
(22)出願日　平成５年(1993)８月19日

(71)出願人　000005108
株式会社日立製作所
東京都千代田区神田駿河台四丁目６番地
(72)発明者　小高　俊之
東京都国分寺市東恋ヶ窪１丁目280番地
株式会社日立製作所中央研究所内
(72)発明者　天野　明雄
東京都国分寺市東恋ヶ窪１丁目280番地
株式会社日立製作所中央研究所内
(74)代理人　弁理士　小川　勝男

(54)【発明の名称】　音声認識装置

(57)【要約】
【構成】　入力手段１、第１の分析手段２、照合手段４、判定手段５よりなる音声認識装置に、入力される音声の様々な様態の変化を検出するための分析を行う第２の分析手段３を設け、その分析結果に基づいて照合手段４あるいは判定手段５を制御する。
【効果】　発声の様態の変化や話者の変化の検出結果に基づいて照合手段あるいは判定手段を制御するので、利用者の多様な発声の様態の変化や話者の変化に対応して音声を認識することができる。

図１



1

【特許請求の範囲】

【請求項1】音声を入力する音声入力手段と、前記音声入力手段により入力された音声を分析し、特徴ベクトルの時系列パタンを出力する第1の分析手段と、予め認識の基準として用意された標準パタンを格納する標準パタン格納手段と、前記標準パタンと前記第1の分析手段から得られる特徴ベクトルの時系列パタンとを照合して、各標準パタンに対するスコアを求める照合手段と、前記各標準パタンに対するスコアに基づいて、一つあるいは複数個の認識候補を出力する判定手段とからなる音声認識装置において、前記音声入力手段により入力された音声に対第2の分析手段を設け、前記第2の分析手段の出力を用いて前記照合手段およびあるいは前記判定手段を制御するようにしたことを特徴とする音声認識装置。

【請求項2】請求項1において、前記第2の分析手段は離散的な値を出力するようにし、前記離散的な値を用いて前記照合手段およびあるいは前記判定手段を制御するようにした音声認識装置。

【請求項3】請求項2において、前記照合手段は前記離散的な値に対応して複数の照合手段を設け、前記離散的な値に基づいて前記複数の照合手段の中から一つあるいは複数個を適宜選択し使い分ける音声認識装置。

【請求項4】請求項2において、前記照合手段は前記離散的な値に対応して複数の照合手段を設け、前記複数の照合手段のすべてあるいは一部を並列動作可能な構成とし、前記離散的な値に基づいて前記複数の照合手段の結果のうち一つあるいは複数個を選択する音声認識装置。

【請求項5】請求項2において、前記照合手段は前記離散的な値に対応して複数の照合手段を設け、前記複数の照合手段のすべてあるいは一部を並列動作可能な構成とし、前記判定手段は前記複数の照合手段から得られる複数の照合結果を、前記離散的な値に基づいて、判定する音声認識装置。

【請求項6】請求項3、4または5において、前記第2の分析手段は、入力される音声の発声単位が音節、単語、文章のいずれであるかを出力する音声認識装置。

【請求項7】請求項3、4または5において、前記第2の分析手段は、話者性に関連した離散的な値を抽出するようにした音声認識装置。

【請求項8】請求項1において、前記第2の分析手段から得られる出力は連続的に変化する量であり、前記連続的に変化する量を用いて前記照合手段および／あるいは前記判定手段を制御するようにした音声認識装置。

【請求項9】請求項8において、前記第2の分析手段は、発声速度に関連した連続的に変化する量を出力する音声認識装置。

【発明の詳細な説明】

【0001】

【産業上の利用分野】本発明は、音声認識装置に係り、特に、同一の話者の発声様態が多様に変化する場合の音

2

声や話者が変わった場合の音声を良好に認識する装置に関する。

【0002】

【従来の技術】従来の音声認識装置、例えば、単語認識装置では、音声を発声する単位が単語であるということを前提としている。この装置に対して複数の単語を続けて発声すると、連続的に発声された複数単語全体が一つの単語であるとみなしてしまい正しい認識結果が得られないことが多い。このように、利用者は単語毎に区切った発声しかできないといった制限を受ける。

【0003】また、音声認識装置が誤認識した場合に利用者が丁寧に一音一音区切って、言い直したりすると、区切って発声された一音一音をそれぞれ一つの単語とみなしてしまい、ますます認識できなくなってしまう。

【0004】

【発明が解決しようとする課題】本発明の目的は、利用者の発声の仕方の変化や話者の変化などにも対応して音声を認識できるようにすることにある。

【0005】

【課題を解決するための手段】上記本発明の目的は、発声の様々な様態の変化や話者の変化の検出を行う第2の分析手段を設け、第2の分析手段の結果に基づいて照合手段あるいは判定手段を制御することにより達成される。

【0006】

【作用】本発明によれば、発声の様態あるいは話者の変化を分析した結果に基づいて照合手段あるいは判定手段を制御するので、発声の多様な様態の変化や話者の変化に対応して音声を認識することができる。

【0007】

【実施例】以下、図を用いて本発明の実施例を説明する。

【0008】図1は本発明の音声認識装置の一実施例を示すブロック図である。本発明で従来と異なっているのは、照合手段4あるいは判定手段5を制御するために第2の分析手段3を設けている点である。入力手段1を通してデジタル化されて入力された音声は第1の分析手段2に送られ、ここで一定時間間隔ごとに音響的な分析が行なわれる。第1の分析手段2の結果は、照合手段4の所望する形式（例えば、特徴ベクトルの時系列パターンあるいはベクトル量子化されたコードの時系列パターンなど）として出力される。照合手段4は、第1の分析手段2から得られる音響的な分析結果である時系列パターンと予め照合の基準として標準パタン格納手段6に用意されている標準パタンとの間で照合を行ない、各標準パタンに対するスコアを出力する。照合手段4から出力されたスコアは、判定手段5に入力され、最もスコアの良い標準パタンに対応した一つあるいは上位の複数の候補が認識結果として出力される。ここまでの入力手段1、第1の分析手段2、照合手段4、判定手段5は従来の音

声認識装置と同様の構成である。本発明で従来と異なっている第２の分析手段３は、入力手段１を通して入力された音声を分析し、発声様態の変化や発声速度の変化を抽出し、この情報を出力する。そしてこの第２の分析手段３の出力により照合手段４あるいは判定手段５を制御する。

【０００９】本実施例では、第２の分析手段で取り出す情報を発声モードとする。発声モードというのは、発声形態、発声様式といった意味のものである。モードといった場合には複数のモードの存在を考えるが、ここでは「音節単位の発声」「単語単位の発声」「文章単位の発声」の三つのモードを考え、それぞれ１）音節モード、２）単語モード、３）文章モードとする。１）の場合は、新しい単語を伝えようとする場合や相手が聞き損なった場合に一音一音丁寧にゆっくりとあるいは区切って発声するような場合であり、例えば、「こ・く・ぶ・ん・じ」と一音一音丁寧に発声する。２）の場合は、コマンドや比較的簡単な情報の伝達を行う場合のように、一つの単語を発声したり、あるいは複数個の単語を単語単位に区切って発声するような場合であり、例えば「国分寺」と発声する。３）の場合は、文章単位でごく普通に発声するような場合であり、例えば、「国分寺まで行きたい」と発声する。

【００１０】次に発声モードを検出する第２の分析手段３について詳しく説明する。

【００１１】図２は発声モードを検出する場合の第２の分析手段の一実施例を示すブロック図である。図３は図２中のブロック図の中で入出力となる情報のいくつかを示しており、(a)～(f)は図２と図３で対応付けられている。図３(a)のような振幅ｗ(ｔ)の音声がパワー算出手段３０１に入力され、図３(b)のようなパワー（短区間パワー）、

【００１２】

【数１】

$$pw(t)=\frac{1}{T}\sum_{i=t}^{t+T} w(t)^2$$

【００１３】が出力される。ただし、Ｔは短区間分析の区間幅である。短区間パワーｐｗ(ｔ)はパワー閾値判定手段３０２に入力され、０（パワー無）／１（パワー有）に変換されて図３(c)のような音声区間ｓｐ(ｔ)が出力される。また、短区間パワーｐｗ(ｔ)はパワー変化量算出手段３０４にも入力され、次式に従って、

【００１４】

【数２】　　ｄｐｗ(ｔ＋１)＝｜ｐｗ(ｔ＋１)－ｐｗ(ｔ)｜

図３(d)のようなパワー変化量ｄｐｗ(ｔ)が算出され

る。パワー変化量ｄｐｗ(ｔ)は、変化量閾値判定手段３０５に入力され、次式に従って、

【００１５】

【数３】　　if　ｄｐｗ(ｔ)≦ＤＰＷTH　then　ｆｉｘ(ｔ)＝１

else　　ｆｉｘ(ｔ)＝０

定常部分かどうか判定され、０（非定常）／１（定常）として図３(e)のように定常区間ｆｉｘ(ｔ)が出力される。ただし、ＤＰＷTHはシステム毎に決められる定数である。次に母音性定常区間判定手段３０６はパワー閾値判定手段３０２からの出力ｓｐ(ｔ)と変化量閾値判定手段３０５からの出力ｆｉｘ(ｔ)を入力として、

【００１６】

【数４】　　ｓｐｆｉｘ(ｔ)＝ｓｐ(ｔ)＆ｆｉｘ(ｔ)（＆は論理積）

により母音による定常区間（母音性定常区間）ｓｐｆｉｘ(ｔ)を図３(f)のように０／１で出力する。続いて定常区間長算出手段３０７は、母音性定常区間判定手段３０６から出力されるｓｐｆｉｘ(ｔ)の０／１の列の中で連続する１の個数により定常区間長（ｆｉｘｓｚ）を求める。定常区間評価手段３０８は、定常区間算出手段３０７により定常区間長が求まる毎に、

【００１７】

【数５】　　if　　ｆｉｘｓｚ≧ＳＺ１TH　then　ｎA＝ｎA＋１

else if　ｆｉｘｓｚ≧ＳＺ２TH　then　ｎB＝ｎB＋１

により、長い定常区間の数ｎA、あるいは短い定常区間の数ｎBを求める。ただし、ｎAとｎBの初期値はともに０である。また、ＳＺ１THとＳＺ２THはシステム毎に決められる定数であり、ＳＺ１TH＞ＳＺ２THである。最後に音声区間検出手段３０３において音声の終端が検出されると、モード判定手段３０９に起動をかける。モード判定手段３０９は、定常区間評価手段３０８よりｎAとｎBを受け取り、以下によりモードを判定する。ここで、ｎは全音節数を表わし、ｎ＝ｎA＋ｎBである。

【００１８】

【数６】　　if　　ｎA／ｎ＞Ｎ１TH

【００１９】

【数７】　　or　ｎ＜Ｎ２TH　　then　モード＝音節モード

else if　ｎ＜Ｎ３TH　　then　モード＝単語モード

else　　モード＝文章モード

ただし、Ｎ１THとＮ２TH、Ｎ３THはシステム毎に決められる定数である。モード判定手段３０９は、まず、全音節数ｎに対する長い定常区間の数ｎAの割合がある閾値を越えているかどうかにより入力された音声がゆっくりと丁寧に発声された音節モードかどうか判定する。さらに、全音節数ｎの大きさによりモードを判定する。このモード判定手段の３０９の出力により照合手段４あるいは判定手段５を制御する。

5

【0020】なお、母音性定常区間を求めるために、ここではパワーの変化だけを用いた実施例を示したが、スペクトルの変化だけあるいはパワーの変化とスペクトルの変化の組合せとしても求められることは言うまでもない。

【0021】次に本実施例の中で用いる照合手段4について図4を用いて説明する。

【0022】図4は、第2の分析手段3の出力を用いてモードを切り替えるようにした場合の照合手段4の構成を示すブロック図である。これは、複数の照合手段の前に選択手段44を設けたものである。選択手段44は第2の分析手段3の出力により複数の照合手段（この例の場合、音節照合手段41、単語照合手段42、文照合手段43）のうち一つあるいは複数（この例の場合は高々二つまで）を適宜選択し、選択された照合手段に第1の分析手段2からの情報を送る。複数選択した場合には判定手段5がスコアに基づいて一つあるいは複数の候補を認識結果として出力することになる。HMM61は、予め統計的に学習された音節単位のモデルを格納している。音節照合手段41はこのモデルに沿って音節単位の照合をし、照合結果として一つあるいは複数の音節の候補をスコアと共に出力する。単語辞書62は、単語についての情報（例えば、どんな音節列で構成されているかに関する情報）を格納している。単語照合手段42は、HMM61に格納された音節単位のモデルを、単語辞書62の情報に沿って組み合わせた単語単位のモデルを用いて単語単位の照合を行い、照合結果として一つあるいは複数の単語の候補をスコアと共に出力する。文法63は、文法を格納している。文照合手段43は、HMM61、単語辞書62、文法63に基づいて照合を行い、照合結果として一つあるいは複数の文あるいは文節の候補をスコアと共に出力する。

【0023】なお、音節照合手段41、単語照合手段42、文照合手段43の実現方法としては様々な方法が考えられるが、ここではHMM（Hidden MarkovModel）を使った方法を考える。HMMを用いた音声認識装置の実現方法については"中川聖一，音声認識における時系列パターン照合アルゴリズムの展開，人工知能学会，Vol.3, No.4, pp414-423, 1988."あるいは"Kai-Fu Lee, Automatic speech recognition: the development of the SPHINX system, Kluwer Academic Publisher, 1989."に詳しく説明されている。

【0024】次に、図5を用いて照合手段4の別の実施例を説明する。

【0025】図5は、第2の分析手段3の出力を用いてモードを切り替えるようにした場合の照合手段4の構成を示すブロック図である。複数の照合手段の後に選択手段44を設けたものである。すなわち、複数の照合手段（音節照合手段41、単語照合手段42、文照合手段43）は並列に動作し、各照合手段からの照合結果のうち

6

一つあるいは複数を、選択手段44が第2の分析手段3の結果に基づいて選択する。音節照合手段41、単語照合手段42、文照合手段43の構成については図4の場合と同じで良い。

【0026】次に、図6を用いて照合手段4のさらに別の実施例を、説明する。

【0027】図6は、第2の分析手段3の出力により判定手段5を制御する場合の照合手段4の構成を示すブロック図である。選択手段がなく、複数の照合結果がすべて判定手段5へ送られる点以外は図4や場合と同じ構成である。

【0028】次に本実施例の中で用いる判定手段5について説明する。

【0029】判定手段5は、入力として照合手段4の出力を受け取る。判定手段5は、最もスコアの良い候補一つあるいは上位の複数の候補を認識結果として出力する。なお、照合手段4内の選択手段44により複数の照合手段が選択されている場合には、それらの照合結果をまとめて、判定手段5がスコアに基づいて最もスコアの良い一つの候補あるいは上位の複数の候補を認識結果として出力することになる。さらに判定手段5は、照合手段4の出力に加えて第2の分析手段3の出力を入力として受け取る場合もある。判定手段5では、第2の分析手段3から受け取った情報（今の場合は、発声モード）に基づいて、照合手段4から送られてきた候補に対してスコアの修正（例えば、重みを付ける）を行ってから、最もスコアの良い一つの候補あるいは上位の複数の候補を認識結果として出力する。

【0030】なお、照合手段4と判定手段5の両方を制御できることは言うまでもない。

【0031】本実施例では、第2の分析手段において発声モードを検出するようにしたが、第2の分析手段が話者性に関連した離散的な値（例えば、男性か女声か、大人か子供か）を抽出するための分析を行なうようにすれば、話者の変化に対応できる。

【0032】また、第2の分析手段が入力音声の発声速度に関連した連続的な値（例えば、音声中の単位時間当りの音節数）を抽出するための分析を行なうようにすれば、発声速度の変化に対応できる。

【0033】

【発明の効果】本発明によれば、発声の様態の変化や話者の変化の検出結果に基づいて照合手段あるいは判定手段を制御するので、利用者の多様な発声の様態の変化や話者の変化に対応して音声を認識することができる。

【図面の簡単な説明】

【図1】本発明の音声認識装置の一実施例を示すブロック図。

【図2】第2の分析手段の一実施例を示すブロック図。

【図3】本実施例の第2の分析手段におけるデータの流れを示す説明図。

7

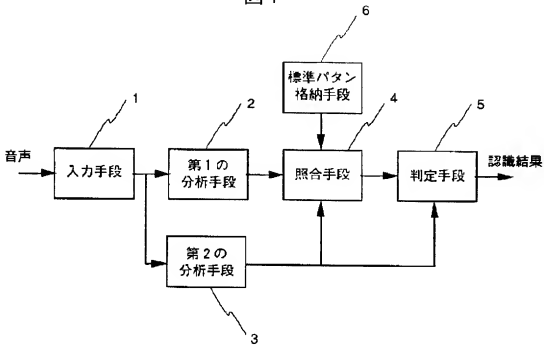
【図４】照合手段の構成を示すブロック図。
【図５】照合手段の他の構成を示すブロック図。
【図６】照合手段のさらに他の構成を示すブロック図。
【符号の説明】

8

１…入力手段、２…第１の分析手段、３…第２の分析手段、４…照合手段、５…判定手段、６…標準パタン格納手段。

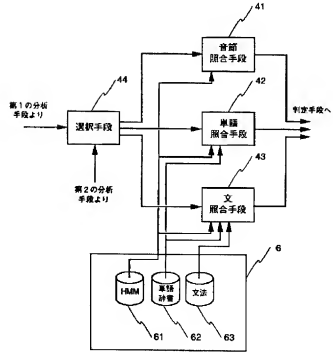
【図１】

図１



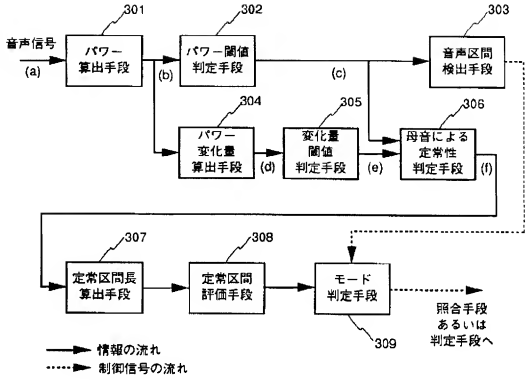
【図４】

図４



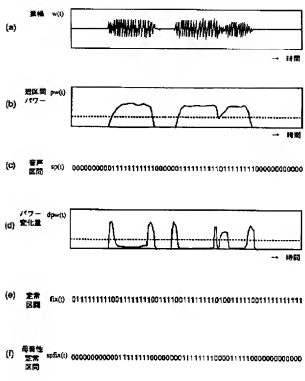
【図2】

図2


音声信号
(a)
301
パワー
算出手段
(b)
302
パワー閾値
判定手段
(c)
303
音声区間
検出手段
304
パワー
変化量
算出手段
(d)
305
変化量
閾値
判定手段
(e)
306
母音による
定常性
判定手段
(f)
307
定常区間長
算出手段
308
定常区間
評価手段
モード
判定手段
309
照合手段
あるいは
判定手段へ
情報の流れ
制御信号の流れ

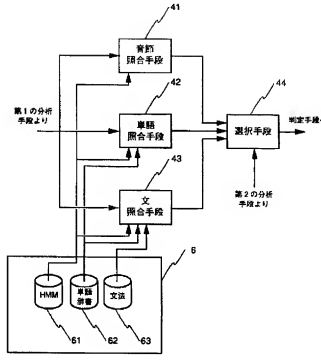
【図3】

図3



【図5】

図5



【図6】

図6